ISSN 2959-6157

The Application of EEG Machine Learning in Schizophrenia

Siyuan Chen

School of CHIPS, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215000, China
Email: csy755935583@gmail.com

Abstract:

Electroencephalographic (EEG) measurements are widely employed in medical and research fields. This review shows the application of EEG machine learning in diagnosing and detecting schizophrenia (SZ), as well as gives the brief background, some diagnostic methods and treatment of SZ with EEG, the latest research findings due to the time of publication, and the advantages and disadvantages of EEG compared to other diagnoses.

Keywords: Electroencephalographic measurements, EEG, schizophrenia, SZ, machine learning, ML

1. Introduction

SZ is one of the costliest mental disorders and typically manifests early, between the ages of 15 and 30. It can lead to cognitive impairments or positive psychotic symptoms such as delusions and hallucinations, as well as the social and economic impact of which on both society and families are substantially negative, and the early detection and treatment of SZ have a positive significance in managing SZ effectively [1]. The methods for diagnosing SZ can be generally divided into two types: the first involves early experiential diagnoses using well-defined criteria, and the second are the scientific diagnoses with equipment. The first diagnostic method is based on the conclusion of experience from lots of authoritative doctors and scholars, so it has a very high accuracy when diagnosing SZ , and importantly, it does not cause any physical harm to the patient. But clearly, this method cannot diagnose and prevent SZ at an early stage, so it may delay the treatment of SZ [2]. The second diagnostic method requires different various equipment. For instance, electroencephalography (EEG), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI) etc. These methods mostly images with the use of penetrating radiation or changes in the body's electromagnetic fields. Moreover, most of them do not invade the human body, ensuring the safety of physical health. Here is an example with MRI. MRI has very high accuracy and could detect changes in the human body at an early stage, which provides a significant advantage in diagnosing diseases and preventing situation becoming worse [3]. However, as MRI technology advances and clinical practices become more complex, safety issues are also increasing. Issues such as projectile forces, risks related to biomedical implants, and device malfunctions are causing injuries to both staff and patients.

EEG, as a non-invasive measurement, detects the electrical signal transmission between brain neurons and translates these signals into an electroencephalogram to analyze brain function and activity states. It ensures physiological safety for the body at a very low cost, making it an ideal method for detecting SZ [4]. When brain cells are activated, they generate local currents. These currents are mainly consisted of Na+, K+, Ca2+, and Cl- ions [5], and conductive electrodes of the EEG measurement attached to the scalp record the changing currents of the brain waves in the form of data. By processing this data, images of brain activity could be generated.

2. Brainwaves Original Signal Groups

Brainwave signals can be divided into four types: δ waves (0.5-4.0 Hz), θ waves (4-8 Hz), α waves (8-13 Hz), and β waves (13-30 Hz) [5]. Electrodes are uniformly distributed across the scalp, as shown in Figure 1, to record the brain signals. Additionally, they could record data in multiple channels to increase accuracy. Among the four waves, α waves are the most common brain waves, and appear when adults are relaxed or have their eyes closed. During activities like thinking or mental calculation, α waves increase significantly. β waves typically appear when the eyes are open, during wakefulness, or in situations of strong mental activity, such as alertness or nervousness. θ waves appear in the beginning of sleep, while δ waves appear in deep sleep.

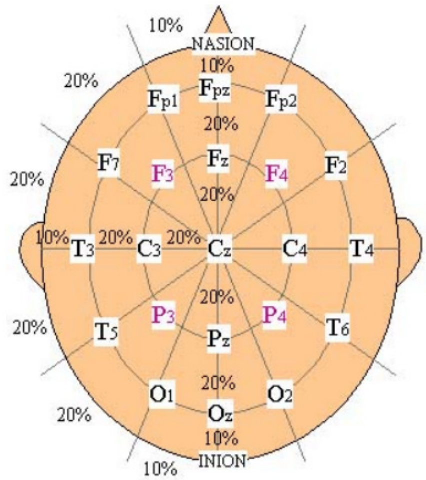

Figure 1 EEG electrode placement on the scalp [5]

3. Machine Learning in EEG

When researchers acquire EEG data from experimental subjects, there are various ways to process it, and machine learning (ML) is one of the most traditional, but also one of the most convenient and efficient methods to utilize. Machine learning is used to quickly analyze data, and through the training process, the machine also becomes more efficient at producing results [6]. In Machine Learning (ML), EEG data firstly undergoes preprocessing to ensure efficient operation on the equipment, followed by analysis using machine learning classification methods such as random forest (RF), logistic regression (LR), support vector machines (SVM), k-nearest neighbors (KNN), and decision trees (DT), among others.

4. Search Method

The study uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [7] guidelines to classify and select studies using the keywords “Electroencephalographic measurements”, “EEG”, “schizophrenia”, “machine learning”, and selects studies related to the topic from valid citation databases including Nature, Google Scholar, Frontiers, IEEE, as shown in Figure 2. The selected 108 Studies range from 2015 to 2023. The author first identifies and removes 16 duplicate Studies by searching for the same keywords. Then, using AI detection and keywords searching methods, 31 completely unrelated Studies are excluded. Due to the initial search scope being too broad, the remaining studies need to be filtered to ensure their relevance to the topic. The first filter, using the keyword “ML” (Machine Learning), successfully excludes 16 studies. Since this study will provide a detailed analysis of datasets, 35 studies that do not mention datasets are filtered out. The remaining 10 studies are highly relevant to the topic and will be the main targets for analyzing.

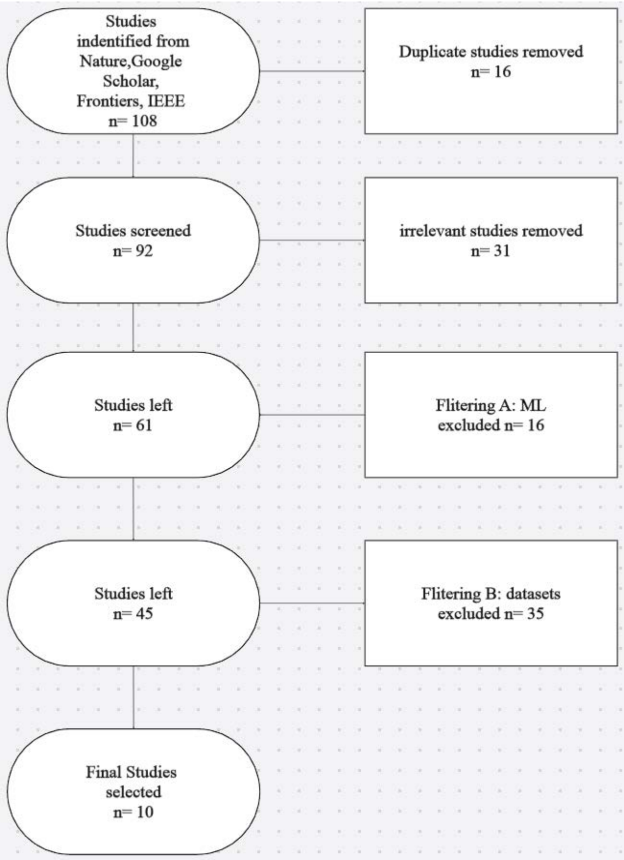

Figure 2 Studies selecting process using PRISMA

5. SZ Diagnoses using Machine Learning

Table 1 summarizes the details from 10 studies, including data preprocessing skills, classification methods, datasets, accuracy and more on.

Table 1 Studies using Machine Learning for SZ diagnoses

Ref	Year	Dataset	Preprocessing	Method	Accuracy
[8]	2020	Institute of Psychiatry and Neurology in Warsaw	Independent Component Analysis (ICA), Dimensionality Reduction ,Spectral Analysis	RF	96.77%
[9]	2022	Institute of Psychiatry and Neurology in Warsaw	Data segmentation, Spectral feature extraction	RF, SVM	RF: 74.99 ± 8.19% SVM: 70.60 ± 5.60%
[10]	2020	Private	-	KNN, SVM, DT	KNN: 44.50% SVM: 43.88% DT: 44.75%
[11]	2021	Private	Wavelet-enhanced independent component analysis (wICA)	Gaussian Naïve Bayes (GNB), Linear Discriminant Analysis (LDA), SVM, KNN, RF	GNB: 96.78% LDA: 96.78% SVM: 98.68% KNN: 91.75% RF: 100%
[12]	2022	Private	EEGLAB, Bandpass filter	Optimized ML (6 features), Optimized ML (19 features)	Optimized ML (6 features): 89.04%, Optimized ML (19 features): 90.93%
[13]	2022	Private	Butterworth filter	RF, The multiple kernel learning (MKL)	RF: 85%,MKL: 83%
[14]	2022	Institute of Psychiatry and Neurology in Warsaw	Bandpass filter, Elimination of bad signal segments, z-score normalization	Recurrent Auto-Encoder (RAE)	81.81%
[15]	2015	FePsy Study	Manual artifact rejection, ICA, Filter, Final manual rejection	Lagged phase synchronisation (LPS), Current-source density (CSD)	LPS: 57%, CSD: 69%, CSD and LPS combined: 70%
[16]	2020	Private	-	SVM, Multinomial naive bayes (Multinomial NB), RF, XGBoost	SVM: 58.2%, Multinomial NB:66.9%, RF: 68.6%, XGBoost: 66.3%
[17]	2021	Private	Second-order Butterworth flter	Adaptive Neuro-Fuzzy Inference System (ANFIS), SVM, CNN	ANFIS: 99.92%, SVM: 92.91%, CNN1: 97.00%, CNN2:98.07%

As shown in Table 1, very high accuracy could be achieved by using data processing method called wCLA, which is larger than 91.75%. The accuracy could even reach 100% when using the RF classification method [11], meaning that complex but systematic data preprocessing steps may help to achieve higher accuracy. Moreover, study [12] used machine learning methods to analyze EEG data by choosing different features, achieving notable accuracy. Study [15] combined LPS and CSD to create a new method, which resulted in higher accuracy with this new approach. These examples reflect the diversity and multiple choices of machine learning in analyzing EEG data. Additionally, two studies [10, 16] mention or use no data preprocessing, resulting in lower accuracy, meaning that data preprocessing plays a quite essential role in EEG data analysis. Moreover, despite the fact that the same data preprocessing methods have been used, the accuracy of different classification methods varies significantly [15, 26]. For instance, in study [16], the accuracy of the RF reached 68.6%, while SVM only achieved 58.2%. This indicates that it is important to choose the appropriate classification method under the same data preprocessing method. Two studies [13, 17] using the Butterworth filter achieve good accuracy, the accuracy reported in study [9] is less certain. In study [12], under the same data preprocessing and classification methods, choosing different numbers and types of features affected the accuracy, indicating that feature selection and optimization are very important in improving accuracy. In these 10 studies, most researchers tend to use private datasets, while some use publicly available datasets from institutions.

The results show that researchers using publicly available datasets provided by institutions tend to achieve very high accuracy. In contrast, studies using private datasets show varying accuracy—some are very high, while others are quite low. This might indicate that public datasets are of high quality, whereas private datasets, due to a lack of standard regulation and careless collection methods, may have quality issues, ultimately affecting the accuracy of the experiments. The results also emphasize the importance of data preprocessing and classification methods in EEG signal processing and highlight the crucial role of balancing these two in machine learning.

6. EEG Datasets for SZ Analysis

From Table 1, it can be concluded that most researchers tend to involve volunteers in data collection experiments with full informed consent in order to collect private datasets [10, 11, 12, 16, 17]. This may be due to the lack of publicly accessible datasets or researchers' doubts about public datasets. The collection of datasets does not follow strict, widely-recognized standards, which leads to both private and some public datasets lacking quality. The root of the problem is that machine learning lacks universality, causing researchers to choose data collection methods based on the data preprocessing and classification methods that they use, that might explain why private datasets are often more relevant to the corresponding study. As a result, private datasets are often only suitable for single studies. However, issues of law and morality limit the circulation of these datasets, making it challenging to build a comprehensive public dataset. Despite this, some institutes and public databases have been providing large amount of publicly available EEG data for diagnosing SZ to deal with the challenge. For instance, some researchers conduct studies using data provided by Institute of Psychiatry and Neurology in Warsaw [8, 9, 14], while others use the database from FePsy Study [15].

7. Challenges in EEG Detection of SZ using ML

In EEG detection on SZ using traditional machine learning, the choice of different databases, data preprocessing, classification methods, and learning algorithms directly impacts accuracy. This makes the application of ML too complex, and a set-up ML environment could not be easily transferred to other studies, indicating that ML's self-learning capability is quite weak. Additionally, due to the privatization of datasets and ethical and legal issues, the construction and development of large public databases face huge challenges. Handling these challenges requires enhancing the open access of data, promoting and resolving ethical and legal concerns, to effectively apply ML in research and practice.

8. Conclusion

This study presents the definition of SZ and some detection methods, the specific principles and advantages of using EEG, as well as the application of ML in EEG data analysis on SZ. While ML can effectively address the issue of SZ detection, it still requires precise control over certain conditions, such as datasets and data preprocessing etc. Additionally, challenges like building large public databases and improving ML's applicability across multiple scenarios still need to be addressed. It is hoped that these issues will be solved in the future.

9. Acknowledgement

This work was supported by Professor Neal Bangerter from Imperial College London.

References

[1] Johannessen, Jan Olav. "Early recognition and intervention: the key to success in the treatment of schizophrenia?" Disease Management and Health Outcomes 9 (2001): 317-327. [5]: Lieberman J A, Dixon L B, Goldman H H. Early detection and intervention in schizophrenia: a new therapeutic model[J]. Jama, 2013, 310(7): 689-690.

[2] Riecher-Rossler, A., Gschwandtner, U., Borgwardt, S., Aston, J., Pfluger, M., Rossler, W., 2006. Early detection and treatment of schizophrenia: how early? Acta Psy-chiatr. Scand. Suppl., 73–80

[3] De Filippis, R., Carbone, E.A., Bruni, A., Pugliese, V., Segura-garcia, C., De Fazio, P., 2019. Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. Neuropsychiatr. Dis. Treat. 15, 1605–1627.

[4] Perrottelli A, Giordano G M, Brando F, et al. EEG-based measures in at-risk mental state and early stages of schizophrenia: a systematic review[J]. Frontiers in psychiatry, 2021, 12: 653642.

[5] Teplan, M., 2002. Fundamentals of EEG measurement. Measurement science review, 2(2), pp.1-11.

[6] Mahesh B. Machine learning algorithms-a review[J]. International Journal of Science and Research (IJSR). [Internet], 2020, 9(1): 381-386.

[7] Sarkis-Onofre R, Catalá-López F, Aromataris E, et al. How to properly use the PRISMA Statement[J]. Systematic Reviews, 2021, 10: 1-3.

[8] Buettner R, Beil D, Scholtz S, et al. Development of

a machine learning based algorithm to accurately detect schizophrenia based on one-minute EEG recordings[J]. 2020.

[9] Ellis C A, Sattiraju A, Miller R, et al. Examining Reproducibility of EEG Schizophrenia Biomarkers Across Explainable Machine Learning Models[J]. 2022.

[10] Ahmedt-Aristizabal D, Fernando T, Denman S, et al. Identification of children at risk of schizophrenia via deep learning and EEG responses[J]. IEEE Journal of biomedical and health informatics, 2020, 25(1): 69-76.

[11] Ciprian C, Masychev K, Ravan M, et al. Diagnosing schizophrenia using effective connectivity of resting-state EEG data[J]. Algorithms, 2021, 14(5): 139.

[12] Keihani A, Sajadi S S, Hasani M, et al. Bayesian Optimization of Machine Learning Classification of Resting-State EEG Microstates in Schizophrenia: A Proof-of-Concept Preliminary Study Based on Secondary Analysis[J]. Brain Sciences (2076-3425), 2022, 12(11).

[13] Santos Febles E, Ontivero Ortega M, Valdés Sosa M, et al. Machine learning techniques for the diagnosis of schizophrenia based on event-related potentials[J]. Frontiers in Neuroinformatics, 2022, 16: 893788.

[14] Wu Y, Xia M, Wang X, et al. Schizophrenia detection based on EEG using recurrent auto-encoder framework[C]// International Conference on Neural Information Processing. Cham: Springer International Publishing, 2022: 62-73.

[15] Ramyead A, Studerus E, Kometer M, et al. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients[J]. The World Journal of Biological Psychiatry, 2016, 17(4): 285-295.

[16] Jo Y T, Joo S W, Shon S H, et al. Diagnosing schizophrenia with network analysis and a machine learning method[J]. International Journal of Methods in Psychiatric Research, 2020, 29(1): e1818.

[17] De Miras J R, Ibáñez-Molina A J, Soriano M F, et al. Schizophrenia classification using machine learning on resting state EEG signal[J]. Biomedical Signal Processing and Control, 2023, 79: 104233.

[18] Najafzadeh H, Esmaeili M, Farhang S, et al. Automatic classification of schizophrenia patients using resting-state EEG signals[J]. Physical and Engineering Sciences in Medicine, 2021, 44(3): 855-870.